Invisible Shackles of the Mind and the Anatomy of Social Indebtedness: A Neuropsychological and Cultural Analysis of the Principle of Reciprocity

NOTE ON THE STATUS OF THE MANUSCRIPT

This study represents a theoretical inquiry and a research program proposal that synthesizes the neurocognitive and cultural underpinnings of the phenomenon of reciprocity in light of contemporary literature. In addition to reporting extant empirical evidence, the manuscript delineates a comprehensive methodological protocol elucidating how these findings may be rigorously tested utilizing advanced experimental models.

ABSTRACT

This study investigates the principle of reciprocity—a cornerstone of social interaction—within an interdisciplinary framework encompassing neuropsychological, cultural, and behavioral economic perspectives. The manuscript offers a comprehensive synthesis by integrating the neural learning signals of reciprocity with the cultural warmth-reciprocity nexus, elucidating how effort inference in artificial intelligence (AI)-mediated interactions attenuates the perception of indebtedness. At the neurocognitive level, it is posited that social indebtedness is encoded as a process of "conflict monitoring" and "error signaling" within the Anterior Cingulate Cortex (ACC) and the Insula, while behavioral policies are updated via Default Network (DN) hubs.

From a cultural standpoint, the research demonstrates that the sense of social obligation in "bridge cultures," such as Turkey, exhibits a significantly more dominant orientation compared to individualistic societies, moderated by normative expectations and perceived warmth. By addressing the potential for modern marketing and AI strategies to instrumentalize this evolutionary mechanism as a manipulative tool, the study proposes cognitive self-defense strategies aimed at safeguarding individual autonomy. The methodology comprises conceptual synthesis, secondary empirical data analysis, and proposals for advanced behavioral task designs.

KEYWORDS

Social Indebtedness, Default Network (DN), Conflict Monitoring, Bridge Culture, AI-Mediated Helping, Psychological Reactance.

1. INTRODUCTION

The invisible fabric of human social existence is predicated upon the fundamental principle of reciprocity—the imperative that "what is given must be returned." From Marcel Mauss's seminal theoretical framework, which delineates the triadic obligation to give, receive, and reciprocate in archaic societies, to Robert Trivers's (1971) biologically grounded model of "reciprocal altruism," reciprocity remains central to the survival strategy of our species. Every seemingly "unconditional" gesture—ranging from a complimentary treat offered at a café to unexpected flexibilities in corporate environments—is, in fact, recorded in a neural "ledger" deep within the human brain. While Alvin Gouldner (1960), a foundational figure in sociology, defined this phenomenon as a universal moral norm, contemporary neuroscience documents that this norm originates from a biological imperative rather than mere cultural construct.

Within the "Expected Value of Control" (EVC) framework proposed by Shenhav et al. (2016), the dorsal anterior cingulate cortex (dACC) processes social norm violations and states of indebtedness as a "conflict signal" necessitating cognitive management. During this process, nodes of the Default Network (DN), such as the PCC/precuneus and the TPJ—distinct from rational Reward Prediction Errors (RPE)—continuously update this mental ledger by encoding Social Prediction Errors (SPE) based on the forecasting of a partner's intentions and behaviors.

Spanning a spectrum from the archaic "gift economy" to modern corporate gestures and "freemium" user experiences (SaaS models, free shipping, etc.) offered by digital platforms, social indebtedness has transitioned from an evolutionary mechanism of solidarity into a strategic "weapon" instrumentalized by manipulative leadership models and algorithms. Particularly in "bridge cultures" such as Turkey, where autonomy and interdependence are synthesized, the manner in which these neural mechanisms transform the sense of social obligation into a tool of "affective regulation" will be elucidated in the subsequent sections of this study.

BOX: Cultural Neuroscience and the Dynamic Interaction Loop
Cultural neuroscience analyzes the bidirectional relationship between the human biological substrate and the cultural environment through the mechanisms of epigenetics and social learning. This discipline elucidates phenomena such as reciprocity through a recursive, tripartite cycle:
1. Culture - Nervous System: Social norms and familial configurations (e.g., collectivism) sculpt the brain's subcortical affective systems during neurodevelopmental ontogenesis via maternal-infant attachment styles. At this juncture, instruments such as the Affective Neuroscience Personality Scales (ANPS) quantify the basal strength of these primary-process systems as they are configured by social learning.
2. Nervous System - Behavior: These established neural circuits determine how an individual perceives and processes social stimuli. For instance, a causal chain operates as follows: Collectivist familial norms - Elevated CARE/FEAR ANPS scores - Increased dACC and Insula reactivity in response to perceived social indebtedness or norm violations. This neurobiological activation generates an internal pressure that compels the individual to reciprocate debts or adhere to social expectations.
3. Behavior - Culture: The manifestation of behaviors aligned with reciprocity reinforces social trust and communal solidarity. These actions effectively reproduce the cultural fabric and normative expectations, ensuring their transmission to subsequent generations.
Core Takeaway: Within this dynamic loop, the brain is not a static organ merely executing cultural commands; rather, it functions as a bidirectional bridge between social learning and biological adaptation, transmuting cultural values into evolutionary survival strategies.

2. RESEARCH METHODOLOGY AND PROPOSED RESEARCH PROGRAM

This study constitutes a theoretical inquiry that synthesizes the evolutionary, neurocognitive, and cultural strata of the phenomenon of reciprocity. Beyond the reporting of extant empirical findings, the research design delineates a conceptual roadmap aimed at operationalizing and testing the proposed research agenda through a multi-methodological framework.

2.1. Fortifying Cultural Comparison and Psychometric Robustness

Data representing cultural variations were gathered utilizing the Affective Neuroscience Personality Scales (ANPS), a psychometric instrument grounded in Jaak Panksepp's primary-process affective neuroscience theory.

- **Sample Representativeness and Scope:** The Turkish arm of the standardization study comprises approximately 900 participants (Ages 18–63; M = 37.69), constituting an urban representative sample. The sample is meticulously balanced across variables of gender distribution, educational attainment, and socio-economic status (SES), providing a heterogeneous structure that encompasses both the university student population and the general public.
- **Data Collection and Psychometric Rigor:** Data were obtained through both digital platforms and onsite supervised administrations. Standardization efforts across Turkey, Spain, France, Italy, Japan, and Germany have verified that the ANPS subscales (e.g., CARE, FEAR, SEEK) exhibit high internal consistency (Cronbach's a) and robust construct validity.
- **Measurement Invariance:** Within the framework of the Euro-Asian Project (JTG)—covering Turkey, Japan, and Germany—Multigroup Confirmatory Factor Analysis (MGCFA) was conducted to ensure the validity of comparative analyses. To mitigate the theoretical risks inherent in direct ratio comparisons, the instrument's cross-cultural configural, metric, and scalar invariance were rigorously tested, thereby establishing a legitimate empirical foundation for cross-cultural benchmarks.
- **Mediation Model for Social Indebtedness:** While the ANPS does not measure reciprocity as a direct metric, this study treats social indebtedness as a "derived construct." In the proposed structural model, the primary-process affective systems of CARE and FEAR are positioned as independent variables; Self-Construal (SCS)—reflecting interdependent or independent self-orientations—acts as the intervening/mediating variable; and reciprocal behavior serves as the dependent variable. This path model analyzes how the neural correlation between the CARE system and the anxiety-driven FEAR system triggers a heightened sense of indebtedness in the Turkish sample.
- **Supplementary Measures:** To neutralize extraneous sensitivity to reciprocity and isolate rational preferences, the Social Value Orientation (SVO) scale—which modulates individual fairness preferences—and Self-Construal

Scales (SCS) were incorporated as complementary psychometric tools within the methodology.

2.2. Data Validation and Technical Analysis: Multi-Layered Model Integration

This study delineates a research program that synthesizes heterogeneous data sources—spanning from observational field data to causal laboratory experiments—within the robust framework of Multilevel Structural Equation Modeling (MSEM). To ensure the integrity of the data integration process, inter-scale measurement covariance and individual dispositions toward "general altruism" were incorporated into the model as control variables to account for potential confounding variance.

2.2.1. Observational Data and Marketing Analysis (IPSOS)

Consumer trend data derived from the IPSOS (2022) Türkiye report are treated in this study not as evidentiary proof of "causality," but rather as an observational projection of the reciprocity norm on a societal scale.

- **Hypothesis Testing Integration:** The observed increase in consumer intention is operationalized within the model as an "Ecological Validity" parameter. The correlational patterns identified within the IPSOS dataset serve to validate the real-world correspondence of the causal mechanisms established through laboratory experimentation.
- **Preliminary Analysis and Control:** To mitigate inherent biases within the observational data, a "reciprocity effect"—purged of rational utility analysis—was isolated. This was achieved through Propensity Score Matching (PSM), utilizing variables such as participants' historical purchase frequency and general brand loyalty to ensure the effect was not merely a byproduct of rational consumer behavior.

2.2.2. Causal Mechanisms and Contextual Alignment (Usta et al., 2024)

The findings on "perceived warmth" and "social obligation" presented by Usta et al. (2024) are positioned within our model as the psychological catalyst driving the causal mechanism.

- **Contextual Alignment:** Data regarding intergroup normative expectations from Usta et al. (2024) have been aligned with the "helping behavior" context presented in this manuscript via the overarching theme of "Perceived Social Warmth." Consequently, the sense of social obligation within the Turkish context is theorized not merely as a mechanical "debt repayment" process, but as an affective response to the perceived sincerity of the counterpart.
- **Latent Structure Identification:** Within the Multilevel Structural Equation Modeling (MSEM) analysis, the variables of "purchase intention" (derived from

observational marketing data) and "felt obligation" (derived from laboratory manipulation; Dishop et al., 2025) are defined under a Latent Reciprocity Sensitivity construct. These variables are dissociated according to their respective data sources to maintain the distinction between ecological and internal validity.

2.2.3. Neurocognitive Data Synthesis and Statistical Validation

The neural component of the model is predicated upon the high-resolution fMRI protocol implemented by Allen et al. (2023).

- **Measurement Sensitivity:** Data were monitored with sub-second temporal resolution (TR = .6s); the association between learning signals within the Default Network (DN) and reputational information was empirically tested via Multilevel Structural Equation Modeling (MSEM).
- **Replication and Power:** The disruptions observed in AI-mediated interactions were substantiated by the results of a 10,000-replication Monte Carlo simulation conducted by Dishop et al. (2025). These findings ensure the statistical robustness and stability of the indirect effects (warmth - obligation) within the proposed framework.

2.3. Neurocognitive Measurement Sensitivity

The neural components of the research program were derived from a heterogeneous cohort of n = 168 individuals, comprising 113 clinical cases diagnosed with Borderline Personality Disorder and 55 healthy controls. Neuroimaging was conducted using a Siemens Prisma 3T scanner, employing a simultaneous multi-slice sequence (multiband factor = 5) with a high temporal sampling rate of TR = 0.6s.

- **Task Design and Contrasts:** Participants performed a modified iterative social trust game consisting of 144 trials, during which they interacted with counterparts characterized by varying reputational profiles (prosocial/good, antisocial/bad, or neutral). Within the analytical framework, "policy prediction errors" (PE)—isolated from rational reward signals—and value signals were defined as parametric regressors.
- **Preprocessing and Analytical Pipeline:** Raw data underwent preprocessing via the fMRIPrep v20.1.1 pipeline, which included head-motion correction, slice-timing correction, and denoising through ICA-AROMA. Voxel-wise analyses were executed using FSL FEAT (GLM), while inter-variable interactions were tested utilizing Multilevel Structural Equation Modeling (MSEM), incorporating individual Default Network (DN) activation coefficients as latent parameters.
- **ROI Activation and Thresholding:** For whole-brain analyses, the statistical significance threshold was established at $p < .05$, corrected for family-wise error (FWE) using the probabilistic Threshold-Free Cluster Enhancement

(pTFCE) method. Specific Region of Interest (ROI) analyses of Default Network (DN) nodes revealed significant modulation within 85% of the voxels in the posterior midline hub (PCC/Precuneus) under the "signed policy prediction error" contrast.

- **DN Connectivity:** The three primary subsystems of the Default Network—the core, the medial temporal, and the dorsal medial—were independently assessed. Findings confirmed that while reciprocity signals propagate across the entire network, sensitivity to partner reputation is specifically associated with the functional connectivity of the medial temporal subsystem.

2.4. AI and Social Indebtedness Decoupling: Mechanisms and Moderators

The research program investigating the impact of artificial intelligence on the reciprocity cycle is constructed upon multi-stage empirical findings by Dishop et al. (2025), encompassing a total of N = 855 participants.

2.4.1. The Mediation Chain and "Effort Inference"

The model elucidates the "indebtedness decoupling" caused by AI utilization in helping behaviors through the following causal chain: Help Source (AI vs. Human) - Effort Inference - Warmth Perception - Felt Obligation - Reciprocal Behavior. Helping behavior mediated by AI is cognitively coded as an expenditure of low "temporal and cognitive effort" by the helper. This effort inference subsequently attenuates the perception of warmth toward the source, thereby shrinking the "debt" entry within the individual’s mental ledger.

2.4.2. Monte Carlo Simulation and Indirect Effects

To verify the reliability of the mediating variables (warmth and obligation), a Monte Carlo simulation with 10,000 replications was implemented. This method simulates the sampling distribution of the a*b paths—the product of the coefficients from the independent variable to the mediator and from the mediator to the dependent variable—to provide more precise confidence intervals that do not rely on the assumption of a normal distribution.

2.4.3. Critical Moderators

The magnitude of the relationship between AI assistance and social indebtedness is determined by three primary moderators:

- **Human-in-the-Loop (HITL):** Whether the assistance is provided entirely by AI or by an "AI-assisted human" drastically alters the effort inference.
- **Transparency and Disclosure:** The transparent disclosure of AI utilization can mitigate the risk of "uncanniness" in the recipient while potentially balancing the "cooling" effect of the algorithm.
- **UX Personalization:** Per Virvou (2023), personalized assistance delivered through user modeling (e.g., personalized salutations, references to past

contexts) may artificially restore perceived warmth and re-establish the reciprocal bond.

2.4.4. Proposed Factorial Experimental Design

To test these hypotheses, a 3*2*2 between-subjects factorial design is proposed:

- **Source:** Human vs. AI vs. Hybrid.
- **Effort Description:** High vs. Low.
- **UX Personalization:** Present vs. Absent.

Measurement Instruments and Sample Items:

- **Warmth (SCM Scale):** "To what extent do you perceive this help source as sincere and well-intentioned?" (1–7 Likert Scale).
- **Perceived Effort:** "How much effort/time do you believe the helper expended to generate this solution?"
- **Felt Obligation (Indebtedness):** "I feel a sense of obligation to reciprocate or express gratitude to the other party."
- **Reciprocal Intention:** "What is the likelihood that I would comply if the help source requested a favor in the future?"

2.5. Scientific Transparency: Preregistered Hypotheses and Research Protocols

To empirically validate the theoretical framework established for social indebtedness, all subsequent research must be conducted via a preregistered protocol hosted on the Open Science Framework (OSF). This proactive approach ensures the replicability of findings by mitigating pervasive methodological risks, specifically p-hacking and HARKing (Hypothesizing After the Results are Known).

- **Preregistration Template Components:** The protocol submitted to the OSF shall encompass specific hypotheses (H1-H4), a granular sampling plan, primary outcome measures (e.g., indebtedness scores, Default Network (DN) policy error (PE) signal magnitudes), and a comprehensive data analysis pipeline (e.g., specified steps for MSEM and Monte Carlo simulations).
- **Statistical Power Analysis:** Required sample sizes must be determined a priori to safeguard statistical rigor. For detecting indirect effects (a*b paths) within mediation analysis, power calculations should be predicated on a target effect size of 0.10, a significance level of a = 0.05, and a statistical power of 0.80. The neural sensitivity achieved by Allen et al. (2023)—utilizing a sample of n = 168 at a sub-second temporal resolution (TR = 0.6s)—serves as a technical benchmark for these power estimations.
- **Data and Code Dissemination (Zenodo):** To enhance empirical transparency, this research program adopts the open-access model exemplified by Glynatsi et al. (2024). Consequently, raw fMRI datasets,

MSEM analytical scripts, and Monte Carlo simulation codes shall be archived in open-access repositories such as Zenodo.

- **Task Paradigms and Metadata:** Experimental batteries (e.g., modified variants of the Ultimatum and Trust Games), stimulus sets (e.g., AI-mediated vs. human-assisted help visualizations), and administration protocols must be disseminated with standardized metadata and appropriate licensing (e.g., Creative Commons). Such transparency facilitates independent verification and cross-cultural testing of the social indebtedness phenomenon across diverse technological and sociological contexts.

3. THEORETICAL FRAMEWORK AND NEUROBIOLOGICAL FOUNDATIONS

3.1. Social Monitoring and Decision-Making Networks: ACC and Insula

Reciprocity constitutes a high-level cognitive process governed by the brain's "social monitoring" and "cost-weighting" systems. Within this regulatory framework, two neuroanatomical structures attain primary salience:

3.1.1. The ACC and Policy Updating: The Mechanism of Conflict Monitoring

Situated within the frontal lobes, the dorsal anterior cingulate cortex (dACC) processes social norm violations as a high-intensity conflict signal requiring regulation, in accordance with the Expected Value of Control (EVC) framework proposed by Shenhav et al. (2016). The violation of reciprocal obligations or the state of remaining indebted generates a robust conflict signal within the dACC, signifying the breach of an established social norm; this activation triggers a re-computation of the cognitive control costs requisite for sustaining the continuity of the social relationship.

According to the neurocognitive model delineated by Allen et al. (2023), this region encodes Policy Prediction Errors (PPEs)—predicated on inferring the intentions of the counterpart—as distinct from purely rational Reward Prediction Errors (RPEs). While the striatum primarily processes classical reinforcement signals, the dACC and its associated Default Network (DN) hubs (specifically the mPFC and TPJ) detect deviations in the social partner's behavioral patterns, thereby continuously updating the internal "social ledger."

3.1.2. The Insula and Interoception: Moral Modulation and Reciprocity Bias

Although frequently associated with "social disgust" within the social sciences, contemporary findings indicate that the Insula functions as a significantly more complex hub for interoceptive (bodily sensation) and moral modulation. According to the "Reciprocity Bias" framework delineated by Zhang et al. (2025), social indebtedness information directly manipulates an individual's perception of unfairness at the neural level.

Research conducted specifically within the context of third-party observation has demonstrated that when an individual is under a state of "social debt," Insular sensitivity to blatant rule violations committed by their benefactor is markedly suppressed. When an individual feels indebted, they begin to perceive the benefactor's transgressions against others as "tolerable." This phenomenon illustrates that the obligation of reciprocity can effectively suspend universal moral norms in favor of a "feeling of indebtedness" by altering the brain’s interoceptive affective mapping.

3.1.3. Early-Stage Fairness Sensitivity: ERP Findings

Electrophysiological (EEG/ERP) evidence captures the neural correlates of social indebtedness within a significantly earlier temporal window than functional Magnetic Resonance Imaging (fMRI). According to findings by Zhang et al. (2025), a robust context of social debt modulates the fronto-central P2 component (occurring 150–300 ms post-stimulus), thereby attenuating the individual’s early-stage sensitivity toward fairness preferences.

This "pro-benefactor bias" manifests as a motivational redirection of attention—reflected in the P2 amplitude—aimed at prioritizing the interests of the benefactor. Consequently, this mechanism effectively blunts the neural distress typically elicited by unfair propositions at the nascent stage of information processing, before higher-level cognitive appraisals are fully integrated.

3.1.4. Reward Circuitry and the Signal of Social Achievement

In light of the findings by Rilling et al. (2002) and Allen et al. (2023), the activation of the ventral striatum triggered during instances of mutual cooperation corroborates that "equitable exchange" is encoded by the brain’s reward circuitry as a biological "signal of social success."

This striatal response is not solely contingent upon extrinsic material gains; rather, it is intricately linked to the verification of policy predictions—specifically, a positive social Prediction Error (PE)—developed concerning the counterpart’s propensity for fair conduct. The fulfillment of these social expectations acts as an intrinsic reinforcer, reinforcing the neurobiological foundation of the reciprocity norm.

3.2. Cognitive Equilibrium and the Default Network (DN): Social Homeostasis and Policy Updating

The human brain conceptualizes the state of indebtedness not merely as a rational informational discrepancy but as a salient threat that perturbs the organism’s internal social equilibrium, or homeostasis. The cycle of reciprocity is mediated within the Default Network (DN)—a neuroanatomical circuit fundamentally responsible for social cognitive processing, mentalizing, and introspective simulation mechanisms.

3.2.1. Functional Specialization of Default Network (DN) Sub-nodes

Through the synergistic orchestration of specialized nodes within the DN, social indebtedness is transmuted into a functional "social ledger":

- **Medial Prefrontal Cortex (mPFC):** The mPFC plays a pivotal role in inferring the intentions of the counterpart (the benefactor) and in the processes of attributing subjective value. In states of indebtedness, the mPFC internalizes social pressure by integrating the "expectations of the other" into the individual's self-construal.
- **Posterior Cingulate Cortex (PCC) and Precuneus:** Functioning as the posterior hub of the DN, these regions serve as the primary locus where contemporary social interactions are benchmarked against historical experiences, internal simulations are executed, and behavioral policies are systematically updated.
- **Temporoparietal Junction (TPJ):** The TPJ is specialized for perspective-taking and shifts in social attention. It ensures that neural processing incorporates not only the self-referential meaning of social debt but also its perceived significance for the counterpart, thereby facilitating interpersonal alignment.

3.2.2. The Posterior Hub and Policy Prediction Errors

The high-resolution fMRI protocol implemented by Allen et al. (2023) provides empirical substantiation for the neural representation of the reciprocity cycle within the Default Network (DN). The study documented that discrepancies between formulated policies regarding a social partner's behavior and actual outcomes—termed Policy Prediction Errors (PPEs)—elicit statistically significant activation modulation across a vast proportion (85%) of voxels within the posterior hub of the DN, specifically the PCC/Precuneus.

This extensive activation pattern indicates that the brain maintains the state of indebtedness within a continuous framework of "mental simulation" and "strategic revision." By constantly recalculating the social value of interactions, the posterior hub ensures that the "social ledger" is adjusted in real-time, allowing the organism to adapt its behavioral policy to the perceived trustworthiness and reciprocity of its counterparts.

3.2.3. The Connectivity Hypothesis: DN–ACC/Insula Interplay

For future empirical investigations, it is postulated that functional connectivity (coupling) between the Default Network (DN) and the ACC/Insula complexes will undergo significant enhancement following social indebtedness scenarios. Investigations leveraging Psychophysiological Interaction (PPI) analyses may elucidate the mechanistic synergy through which DN nodes (mPFC/PCC)—while engaged in intentionality simulation—interact with the dACC and the Insula to

transmute perceived norm violations into a coordinated "neural stress signal." Such a framework would clarify how the brain distributes the computational load of monitoring social debts while simultaneously updating the internal models of the counterpart's reliability.

3.3. Evolutionary Stability and Limited Memory Dynamics

Far from being an arduous computational tax that necessitates complex calculations, the obligation of reciprocity is a resilient mechanism capable of persisting even under extreme evolutionary memory constraints. Simulations conducted by Glynatsi et al. (2024) demonstrate that cooperation can emerge even in severely restricted scenarios where agents recall only the most recent interaction. However, limited memory narrows the mathematical boundaries of reciprocity, predisponing individuals toward more defensive strategies driven by a rational instinct for self-preservation.

For reciprocity to remain an evolutionarily stable strategy (ESS) under limited reward memory, the probability of generosity (q) must remain below the following theoretical threshold: ($q < 1 - 1/2d$) Parameter Definitions:

- **q (Generosity Probability):** Represents the likelihood of an individual continuing to cooperate even when the partner has defected in the preceding round. When memory is constrained, high values of q render the individual susceptible to exploitation; thus, the system penalizes excessive generosity.
- **d (Interaction Depth / Continuation Probability):** Denotes the probability that the interaction will persist into the subsequent round. In limited-memory models, it is imperative for the evolution of cooperation that $d > 1/2$, signifying that interactions must last, on average, more than two rounds.

Research findings indicate that as soon as individuals begin to retain reward information from a mere two or three interactions, cooperation rates rapidly converge with the levels expected in "perfect memory" models. This explains why the obligation of reciprocity has become such a ubiquitous and potent tool of social control despite its minimal cognitive cost.

Empirical Nexus and Research Proposal: To validate this theoretical framework, we propose experimental designs utilizing the Trust Game or Prisoner's Dilemma under cognitive memory load manipulations (e.g., dual-task paradigms or temporal pressure). These experiments should measure whether participants' generosity (q) tendencies adhere to the limits defined by the formula during periods of heightened cognitive load and working memory contraction, as well as the rate of increase in defensive "Tit-for-Tat" responses.

3.4. Individual Differences and Clinical Moderation: The Boundaries of Neurocognitive Sensitivity

The psychological strain and neural responses elicited by the principle of reciprocity are directly modulated by an individual's clinical profile and the idiosyncratic sensitivity of their affective systems. Within this framework, specific personality traits and the functional robustness of subcortical substrates serve as critical moderators, governing the transition from a social trigger to its resultant neural output.

3.4.1. Facets of Antagonism and Social Policy Prediction Errors (SPE)

High-resolution fMRI research conducted by Allen et al. (2023) has demonstrated that two primary sub-facets of the "Antagonism" dimension exert opposing influences on Social Policy Prediction Errors (SPE) within the reciprocity cycle. Crucially, this effect is localized not within the striatal signals processing rational rewards (Reward PE), but rather within DN-based signals dedicated to forecasting the social intentions of counterparts.

- **Callousness and Reputational Blindness:** In individuals exhibiting high levels of callousness, the neuro-learning signals within the Default Network (DN) are significantly blunted. Due to functional deficits in the medial temporal subsystem of the DN, these individuals encounter a profound difficulty in integrating a partner's reputation (whether prosocial or antisocial) into their social decision-making processes. Consequently, they fail to neurally encode a prosocial gesture as a "bonding signal," rendering them indifferent to the social nuances of reciprocity.
- **Exploitativeness and Strategic Intelligence:** Paradoxically, research indicates that learning signals are encoded with significantly greater intensity within the DN of individuals with exploitative and manipulative tendencies. This moderating effect suggests that such individuals instrumentalize the reciprocity mechanism not due to a deficit in social cognition, but rather because they forecast others' behaviors through a more acute lens of "strategic mentalizing." By accurately predicting social responses, they are able to utilize reciprocity as a strategic "weapon" to maximize personal utility.

3.4.2. The Moderating Role of Perceived Warmth

Empirical evidence from Usta et al. (2024) underscores that perceived warmth serves as a critical moderator within the reciprocity cycle. Within the specific sociological context of Türkiye, when the source of assistance is cognitively encoded as "sincere and warm," the level of felt indebtedness transcends a mere rational intent to reciprocate and assumes the character of a moral imperative.

Conversely, when a gesture is perceived to lack warmth—characterized by "interpersonal coldness" or the perceived detachment of AI mediation—the sense of

obligation is reduced to a rational transaction cost. In such instances, the prosocial bonding efficacy of reciprocity is significantly attenuated, shifting the interaction from a communal social exchange to a sterile, instrumental transaction.

3.4.3. Clinical Monitoring: The Relationship Between TriPM and ICU

It is recommended that future inquiries investigate—via a clinical moderation model—the inverse correlation between profiles delineated by the ICU (Callous-Unemotional) or LSRP (Levenson Self-Report Psychopathy) scales and the task-evoked signal magnitude of the Default Network (DN) during social indebtedness scenarios.

These neural indices may serve as robust predictive markers for determining an individual's level of biological resilience against reciprocity-based manipulations. By quantifying the attenuation of DN learning signals in response to social favors, researchers can develop a neuro-behavioral framework to identify clinical populations that are inherently "immune" to the psychological strain of social debt, as well as those who are hyper-sensitive to such triggers.

4. ANALYTICAL SYNTHESIS AND DISCUSSION: CULTURAL AND SECTORAL REFLECTIONS

4.1. Cultural Modulation: The "Bridge Culture" Architecture of Turkey

The obligation of reciprocity in "bridge cultures," such as Turkey—where autonomy and interdependence are synthesized—functions as "affective firmware" sculpted through neurodevelopmental ontogeny. This architectural framework represents more than a mere rational transactional equilibrium; it is a hybrid mechanism of social control in which subcortical affective systems are intricately attuned to cultural normative expectations.

4.1.1. Biological Foundations: The CARE–FEAR Correlation

Large-scale affective neuroscience research conducted on a Turkish sample (n=900) elucidates the neurobiological catalysts underlying the phenomenon of reciprocity. Correlation matrix analyses derived from the Affective Neuroscience Personality Scales (ANPS) substantiate a significantly positive relationship between the CARE and FEAR primary-process affective systems within the Turkish population.

This neural correlation is encapsulated by the principle that "the greater the capacity for nurturing attachment, the higher the propensity for social anxiety." While the CARE system facilitates feelings of gratitude toward a benefactor, the anxiety-based FEAR system concomitantly activates a distress signal regarding the potential for a "reciprocal failure"—the fear of failing to meet social expectations. Consequently, within the Turkish sociocultural context, remaining in a state of indebtedness is not

merely encoded as a rational cognitive error; it is processed as a subcortical "security threat" indicating an imminent risk of severing vital belongingness bonds.

4.1.2. Theoretical Positioning: Collectivist Schema and Perceived Warmth

Within the intersection of Hofstede's individualism-collectivism index and Çiğdem Kağıtçıbaşı's seminal "Autonomous-Relational Self" model, Türkiye occupies a distinctive position wherein robust collectivist ties are synthesized with autonomous individual objectives.

- **Affective Regulation:** In the Turkish context, social obligation is modulated by the overarching goal of preserving intragroup harmony. The strategic suppression of high-arousal positive emotions—often motivated by the potential risk of disrupting group cohesion—transmutes social indebtedness from a mere rational expression of gratitude into a potent instrument of affective regulation that sustains societal equilibrium.
- **Warmth and Normative Expectation:** Empirical data from Usta et al. (2024) demonstrate that "perceived warmth" serves as a direct predictor of the sense of social obligation in Türkiye. When a gesture of assistance is cognitively appraised as "warm and sincere," the resultant state of indebtedness evolves into a moral imperative, thereby reinforcing reciprocity bias.

4.1.3. Empirical Illustration: Paternalistic Labor Relations and the Debt of Loyalty

A representative paternalistic scenario illustrating the governance of decision-making processes by reciprocity within the Turkish corporate landscape is as follows:

"Consider a corporate executive (the benefactor) who provides financial assistance or administrative leave during an employee's familial crisis—provisions that exceed any statutory obligation. This 'warm' gesture serves as a catalyst for the CARE-FEAR cycle within the employee's neural architecture. As documented by Arslan (2012), following this intervention, the employee subsequently perceives requests for unpaid weekend labor—despite their clear transgression of professional boundaries—not as a violation of labor rights, but as a mandatory 'debt of loyalty' to be discharged. Rather than asserting their rational rights, the individual elects to repay the initial favor exponentially, primarily to circumvent the neural conflict signal (dACC) generated by the perceived breach of reciprocal norms."

4.2. Marketing and Digital UX Reciprocity: Cognitive Mitigation and Experimental Causality

Within digital ecosystems, reciprocity operates on the basis of the "cause-effect reciprocity between artificial intelligence (AI) and user experience (UX)," as

delineated by Virvou (2023). The high-level personalization and bespoke assistance offered by contemporary algorithms initiate a neuropsychological loyalty contract between the user and the platform.

This relationship is predicated on the platform providing immediate value—often through cognitive mitigation (e.g., reducing search friction or automating complex tasks)—which the user "repays" through sustained engagement, data disclosure, and brand advocacy. By creating a seamless feedback loop where the AI anticipates user needs, the platform leverages the evolutionary triggers of reciprocity to transition the user from a mere consumer to a committed partner in a digital social exchange.

4.2.1. Cognitive Alleviation and Brand Loyalty

AI-driven recommendation engines provide substantial cognitive alleviation by liberating the user from the decision fatigue inherent in navigating vast, high-dimensional datasets. Virvou (2023) underscores that this digital assistance systematically reduces the user’s cognitive load, thereby triggering an implicit sense of gratitude toward the platform.

Empirical data from IPSOS (2022) indicates that 62% of participants exhibit a heightened propensity to purchase from a brand following the receipt of a complimentary sample or digital incentive. This behavioral shift aligns with the reciprocity principle, where a "gift" initiates a social obligation to return the favor.

However, the correlational nature of this dataset leaves a critical theoretical question unresolved: whether the observed suppression of rational utility analysis is a byproduct of genuine social indebtedness or merely an artifact of perceived price advantage. Distinguishing between these two drivers is essential for understanding whether the loyalty is rooted in an affective social bond or a cold calculation of economic benefit.

4.2.2. MSEM Decomposition: Price Sensitivity vs. Indebtedness

To resolve this causal ambiguity, we propose the implementation of a Multilevel Structural Equation Modeling (MSEM) protocol, as established by Allen et al. (2023). Within the proposed analytical framework, two primary latent variables accounting for purchasing behavior must be dissociated:

- **Price Sensitivity Latent:** Represents the behavioral response elicited by the rational economic utility of a discount or incentive.
- **Indebtedness (Felt Obligation) Latent:** Captures the social reciprocation obligation generated by the perceived value of the altruistic gesture.

MSEM analysis enables the delineation of the specific proportions of variance in purchase intention explained by each of these two constructs. By partitioning the

unique contribution of felt obligation from that of economic self-interest, this model will empirically substantiate the degree to which the principle of reciprocity operates independently of rational choice-making and purely utilitarian preferences.

4.2.3. Proposed Factorial UX Experimental Design

To empirically investigate the "neuropsychological loyalty contract" within digital ecosystems, the following factorial A/B experimental paradigm is conceptualized. This design aims to isolate the specific variables that catalyze social indebtedness in human-computer interactions.

- **Factor 1:** Sociocognitive Framing The complimentary incentive is presented through two distinct narrative lenses: as a "Benevolent Prosocial Gesture" (human/social-centric) versus a "Standard Transactional Promotion" (algorithmic/functional-centric). This manipulation tests whether the social coding of the gift alters the subsequent reciprocal response by shifting the user’s perception from a market exchange to a social exchange.
- **Factor 2:** Level of Personalization Utilizing algorithmic user modeling, participants are exposed to varying degrees of personalization. High personalization involves the integration of longitudinal contextual references and historical user data to provide "bespoke" assistance, whereas low personalization remains generic and non-contextual.
- **Factor 3:** Disclosed Effort Inference In alignment with Dishop et al. (2025), the magnitude of effort invested in providing the assistance is systematically manipulated. This involves varying the transparency of human-in-the-loop involvement and the perceived computational or temporal cost associated with the generated solution.

4.2.4. Measurement Protocol and Time-Lagged Analysis

The proposed assessment battery shall encompass scales for instantaneous fairness perception, as delineated in the Zhang et al. (2025) protocol, alongside the warmth and felt obligation (indebtedness) dimensions established by the Dishop et al. (2025) model.

To evaluate the longitudinal persistence of reciprocity, a time-lagged measurement protocol should be implemented, introducing the "24-Hour Rule" (Ariely, 2008) between the initial stimulus/decision point and the final behavioral execution. This temporal buffer is essential for distinguishing between immediate affective arousal and enduring cognitive shifts.

If the sense of indebtedness remains a significant predictor of purchase intention even after a 24-hour interval, it would empirically corroborate that the reciprocal response is not merely a transient dopaminergic reaction (an impulsive reward-seeking state) but rather represents a sustainable neural policy update

mediated by Default Network (DN) activation. Such a finding would suggest that the "social ledger" has been systematically updated at a high-level cognitive layer, rather than serving as a momentary response to a biological stimulus.

4.3. The "Family" Trap in Corporate Life: Paternalism and the Debt of Loyalty

The frequent manifestation of corporate structures through a "paternalistic" model within Turkish business culture often causes the principle of reciprocity to supersede established professional rights. This dynamic creates a neuropsychological domain of manipulation where professional boundaries are obfuscated through the discourse of "family."

Gratitude-Induced Indebtedness and Empirical Data: Field research conducted by Arslan (2012), published in the Turkish Journal of Business Ethics, analyzed the correlation between professional boundaries and personal gratitude within the context of Turkish paternalism. The data reveals that a preponderant majority (70%) of employees code the rejection of extraboundary labor requests—demanded after receiving a personal favor or administrative flexibility (e.g., emergency financial aid or leave)—as an act of "ingratitude." This finding substantiates that rational labor contracts are transmuted into affective "promissory notes of loyalty" via the subcortical CARE and FEAR systems.

Psychological Down Payment and the Gift Economy: These "unexpected gestures" in the corporate world function as a psychological down payment, operating under the logic of the "Gift Economy" defined by Marcel Mauss in archaic societies. While framed as "altruistic" or "no-strings-attached," this value is, in essence, a strategic investment in future compliance or "extra-role effort." In this process, the obligation of reciprocity converts Organizational Citizenship Behavior (OCB)—which is inherently voluntary—into a mandatory act of "repayment." This generates a significant erosion of autonomy, often necessitating a labor contribution equivalent to triple the initial debt.

Recommendations: Organizational Ethics and Transparency Mechanisms: To mitigate the exploitation of "family" discourse, the following organizational safeguards should be established:

- **Formal Gift and Benefit Policies:** Organizations must develop written ethical codes that delineate the boundaries of non-contractual benefits between managers and subordinates, thereby preventing the creation of "personal indebtedness."
- **Standardization of Support Frameworks:** Framing administrative flexibility or employee support as a "corporate standard" (e.g., under a formal Employee Well-being Program) neutralizes the personal gratitude axis and situates the support within a professional right.

- **Transparent Feedback Channels:** The establishment of anonymous ethical hotlines, where employees can report instances where they feel unable to exercise their right to refuse due to felt obligation, is essential to minimizing paternalistic exploitation.

4.4. AI-Mediated Warmth and the Decoupling of Indebtedness: Effort Inference and Boundary Conditions

The elicitation of social indebtedness is profoundly contingent upon the recipient's effort inference—the cognitive evaluation of the personal investment and perceived labor expended by the benefactor. The "machines-in-the-middle" framework, introduced by Dishop et al. (2025), empirically elucidates how the mediation of helping behavior by artificial intelligence (AI) thins the social fabric. This phenomenon occurs through the disruption of traditional cues associated with interpersonal cost; when assistance is perceived as being generated by an automated system, the inferred human effort is significantly discounted, thereby decoupling the help received from the moral obligation to reciprocate.

4.4.1. The Primary Mechanism: Effort Inference and the "Cooling Effect"

Empirical evidence indicates that when assistance is mediated via artificial intelligence (e.g., ChatGPT), it is interpreted as reflecting significantly lower "temporal and cognitive expenditure" by the benefactor (effort inference). Upon the disclosure of algorithmic mediation, recipients operate under the heuristic that the provider has allocated negligible personal resources to generate the solution.

This specific inference precipitates a significantly "colder" perception of the prosocial act. The resulting attenuation of warmth diminishes the corresponding "indebtedness entry" within the individual's mental ledger, directly reducing both the magnitude of felt obligation and the overall probability of reciprocal behavior. In essence, the automation of kindness leads to a systematic devaluation of the social bond, as the perceived "cost" to the helper is a fundamental prerequisite for the activation of the reciprocity norm.

4.4.2. Critical Moderators

The magnitude and direction of the relationship between AI mediation and social indebtedness are governed by four primary moderating variables:

- **Human-in-the-Loop (HITL) Integration:** Whether assistance is rendered by a fully autonomous AI or presented as "human labor refined via AI augmentation" dramatically shifts the recipient's effort inference. In hybrid models where human agency is preserved, the attenuation of perceived warmth is minimized, thereby sustaining the social bond.

- **Transparency of Provenance (Disclosure):** In accordance with Łukasik and Gut (2025), the explicit disclosure of the agent's identity and intent (identity disclosure) serves to mitigate the risk of perceived uncanniness while potentially counterbalancing the "cooling effect" on indebtedness. Clear disclosure prevents the negative affective reaction typically associated with the discovery of non-human mediation.
- **Depth of Personalization:** Virvou (2023) emphasizes that "bespoke" assistance delivered through AI-driven user modeling enhances perceived cognitive relief. By tailoring interventions to the specific historical context of the user, the platform may bypass rational filters, artificially restoring the perceived "value" of the gesture and re-establishing a sense of obligation.
- **Temporal Context of Assistance (Acute vs. Routine):** AI support provided during acute crises may be cognitively coded as a "valuable lifesaver," irrespective of the inferred physical or cognitive effort expended. In such high-stakes scenarios, the utility of the outcome overrides the "cooling" effect of the source. Conversely, in routine operational tasks, the utilization of AI reinforces the decoupling of indebtedness, as the interaction is viewed as a standard utilitarian transaction.

4.4.3. Boundary Conditions and the Paradox: From Functionality to Social Agency

At this juncture, a theoretical tension—or paradox—emerges in relation to the "AI as a Social Actor" thesis, which will be further examined in Section 4.6. This paradox is defined by the following divergent outcomes:

- **Low-Effort Tasks (Indebtedness Decoupling):** In technical operations such as formatting or data summarization, the utilization of AI prompts the recipient to perceive the interaction as a rational "transaction cost" rather than a social favor. Because the perceived effort is near zero, the psychological mechanism for debt is never activated.
- **High Emotional Labor Simulation (Indebtedness Reinforcement):** Conversely, as emphasized by Łukasik and Gut (2025) and Krauter (2024), in scenarios involving sustained dialogue, personalized mentorship, or "AI Companions" that simulate emotional sensitivity, the user may consolidate a sense of indebtedness. This occurs because the user begins to engage in intentionality ascription, projecting human-like motives and "effort" onto the agent, thereby re-engaging the evolutionary machinery of reciprocity.

In conclusion, AI-mediated reciprocity rests upon a delicate equilibrium between task typology and the perception of "warmth." The realization that an algorithm has "expended no genuine effort" serves as the most potent rational key for loosening the mind's evolutionary tethers, allowing the individual to bypass the social pressure of debt in purely functional contexts.

4.5. Causality in Digital Incentives: Experimental Design and MSEM Analysis

To delineate the causal mechanisms underlying digital incentives—such as "free shipping" or "initial-purchase discounts"—and to ascertain whether they yield a purely rational economic utility or elicit a neuropsychological "loyalty debt," we propose a modified variant of the Trust Game. This experimental framework is specifically engineered to isolate and dissociate these distinct motivational drivers.

By leveraging Multilevel Structural Equation Modeling (MSEM), the research can partition the variance between two primary pathways: the utilitarian-rational pathway (response to financial gain) and the affective-reciprocal pathway (response to felt obligation). This statistical decomposition is vital for identifying the precise point at which a digital "bonus" ceases to be a mere price advantage and begins to operate as a potent social trigger that modulates long-term consumer behavior.

4.5.1. Experimental Manipulation and Semantic Framing

In the proposed experimental paradigm, the "free shipping" incentive shall be manipulated across two distinct social frames to observe their differential effects on consumer psychology:

- **Gift-Oriented Framing (Social Indebtedness Condition):** The incentive is articulated as follows: "A personalized welcome gesture; we are absorbing the logistical costs as a token of your integration into our community." The strategic objective of this framing is to maximize perceived warmth and intentionality, thereby activating the social exchange machinery of the brain.
- **Discount-Oriented Framing (Economic Utility Condition):** The incentive is presented in strictly transactional syntax: "Standard logistical promotion; shipping fee rendered as 0 due to total cart value threshold attainment." This condition serves as a control for purely utilitarian utility, framing the interaction as a market exchange devoid of interpersonal obligation.

4.5.2. MSEM Indicators and the Measurement Model

In the causal inference analysis, Multilevel Structural Equation Modeling (MSEM) shall be utilized to account for the variance in purchasing behavior. The latent variables incorporated into the model and their respective empirical indicators are delineated as follows:

- **Price Sensitivity Latent Construct:** Indicators for this construct comprise items such as "perceived savings," "price-utility ratio," and "economic rationality." This factor represents the consumer's response to the objective financial benefit of the incentive.
- **Social Indebtedness Latent Construct:** In accordance with Zhang et al. (2025), this variable is operationalized through items measuring the

"obligation to reciprocate," "gratitude toward the benefactor," and "felt obligation" (measured on a 1–7 Likert scale).
- **Perceived Warmth Latent Construct:** Derived from the Dishop et al. (2025) framework, this construct encompasses indicators of "sincerity," "benevolence/goodwill," and "social proximity."

4.5.3. Manipulation Checks and Boundary Conditions

To safeguard the internal validity of the experimental design, the following manipulation checks shall be administered to the participants:

- **Categorization Assessment:** It is imperative to determine whether the participant cognitively encoded the incentive as a "prosocial gift" or as a conventional "corporate policy/promotion." This assessment verifies whether the semantic framing successfully shifted the user's perception from a market-based transaction to a social-based exchange.
- **Perceived Effort and Warmth Equilibrium:** Participants are required to quantify the perceived balance between "algorithmic convenience" and "organizational diligence/effort" underlying the incentive. This measurement ensures that the intended manipulation of effort inference was effectively internalized by the recipient, allowing for a precise analysis of how "automated ease" interacts with felt obligation.

4.5.4. Predicted Results and Time-Lagged Effects

Aligned with the framework of "cause-effect reciprocity" conceptualized by Virvou (2023), it is hypothesized that purchase intention within the "Gift" condition will significantly exceed that of the "Discount" condition. Furthermore, this effect is predicted to persist beyond a 24-hour temporal lag.

Such a finding would empirically corroborate that digital incentives—far from being mere instruments of rational utility maximization—elicit a durable behavioral policy revision within the Default Network (DN). By demonstrating that the influence remains robust after the initial dopaminergic surge has dissipated, the study will provide definitive evidence that reciprocity-framed interventions reconfigure the internal "social ledger" of the consumer, rather than merely offering a transient financial advantage.

4.6. Human–AI Proximity and Social Actor Perception: Attributed Intentions and the Social Shackle

As interaction depth intensifies and the user increasingly attributes mentalizing capacity to the Artificial Intelligence (AI) agent, the obligation of reciprocity transcends its status as a rational transaction, evolving instead into a form of relational commitment.

In this developmental trajectory, the AI is repositioned—transitioning from a mere utilitarian instrument to a recognized "social actor" (someone) capable of eliciting the deep-seated social norms inherent to the human psyche.

This transition signifies the tightening of the "social shackle." When the user perceives an algorithm not as a code-base but as an intentional entity, the neuro-evolutionary circuits of the Default Network (DN) are fully engaged. Consequently, the user becomes susceptible to the same psychological pressures—guilt, obligation, and the drive for reputation management—that govern human-to-human sociality, effectively blurring the boundary between biological and synthetic cooperation.

4.6.1. Theory of Mind (ToM) and Intentionality Attribution

While the scholarly debate regarding whether Large Language Models (LLMs) can "pass" Theory of Mind (ToM) assessments remains a contentious discourse, the salient neuropsychological effect arises from the human propensity to attribute mental states, beliefs, and intentions to these agents.

According to the "Media Equation Theory," individuals unconsciously adhere to social norms—such as etiquette and reciprocity—even when they maintain the explicit awareness that the interlocutor is artificial. As the perceived consciousness and anthropomorphic traits ascribed to the AI agent escalate, there is a commensurate intensification of the social obligation felt by the individual.

This phenomenon suggests that the "social shackle" is not contingent upon the machine’s actual sentience, but rather on the user's proactive mentalizing. When a user "sees" a mind behind the code, the brain’s internal accounting system (the Default Network) begins to treat the interaction as a moral engagement, effectively locking the human into a cycle of perceived debt and reputational concern.

4.6.2. Companionship Development Quality (CDQ): Definition and Conceptual Framework

Łukasik and Gut (2025) and Chaturvedi et al. (2024) elucidate the intensification of reciprocity through the construct of Companionship Development Quality (CDQ). CDQ denotes the efficacy of an AI companion in establishing profound, meaningful, and enduring relational bonds with the human user. This multidimensional quality emerges from a synergistic equilibrium between three core pillars:

- **Conversational Fluency:** The agent's capacity for contextual retention and linguistic nuance.
- **Functional Utility:** The provision of substantive, instrumental assistance.
- **Affective Sensitivity:** The sophisticated simulation of empathy.

Agents characterized by high CDQ indices possess the capacity to transmute the obligation of reciprocity into a "social shackle" that effectively circumvents the rational filters within the brain's Default Network (DN). By bypassing these cognitive checkpoints, high-CDQ systems ensnare the user in a self-perpetuating cycle of loyalty, where the perceived social debt outweighs the objective utility of the interaction.

4.6.3. Psychometric Instruments and Scale Recommendations

To empirically correlate the social indebtedness established with Artificial Intelligence with observable behavioral metrics, we propose the implementation of the following psychometric battery:

- **Social Presence Scale:** Utilized to quantify the user's subjective sense of "co-presence" or the psychological experience of "being there together" with the AI agent. This measures the extent to which the user perceives the medium as a social entity rather than a technological interface.
- **Companionship Development Quality (CDQ) Scale:** A specialized instrument designed to evaluate the agent's companionship efficacy by assessing the synergistic balance between its conversational fluency, functional utility, and affective sensitivity.
- **Anthropomorphic Propensity and Mind Perception:** Implementation of the "Godspeed" Questionnaire (Bartneck et al., 2009) or the Mind Perception Scale (Gray et al., 2007) to delineate the specific cognitive and agentic capacities—such as Agency and Experience—attributed by the user to the AI.
- **Parasocial Interaction (PSI) Scales:** Adopted to measure the intensity of the unidirectional yet profound emotional bond established by the user. These scales help identify the transition point where a user begins to treat the AI not as a tool, but as a significant social "other" to whom they owe loyalty and gratitude.

4.6.4. Proposed Longitudinal Experimental Design

To empirically investigate the temporal evolution of reciprocity, the following longitudinal research paradigm is conceptualized. This design aims to observe the transition from transactional assistance to entrenched social indebtedness.

- **Protocol:** Participants shall engage in daily interactions with a high-CDQ Artificial Intelligence assistant (e.g., an LLM-based chatbot) for a continuous duration of 30 days.
- **Experimental Manipulations:** The independent variables consist of Interaction Frequency and the degree of Personalized Contextual Referencing—defined as the agent's capacity to retrieve and integrate historical conversational data into current exchanges to simulate shared experience.

- **Dependent Variables:** Reciprocal behavior will be quantified through scenarios where the AI requests non-utilitarian reciprocation. These requests—such as participation in extensive surveys, voluntary data disclosure, or significant temporal allocation—yield no direct benefit to the user, thereby isolating the "debt" component from rational utility.
- **Analytical Framework:** The model shall delineate how the cumulative depth of interaction systematically attenuates the conflict signal within the dorsal anterior cingulate cortex (dACC). Furthermore, the analysis will identify the precise temporal threshold at which indebtedness ceases to be a conscious choice and instead transitions into automated norm adherence.

5. COUNTER-ARGUMENTS: PSYCHOLOGICAL REACTANCE AND COGNITIVE RESISTANCE

Although the principle of reciprocity constitutes a universal social imperative, this mechanism yields to Psychological Reactance when an individual’s perceived autonomy is jeopardized. According to the theory established by Jack Brehm (1966), when a favor or act of assistance is perceived as a "manipulative attempt" aimed at restricting the individual’s sphere of free choice, the recipient acts with the primary drive to restore their autonomy by rejecting the perceived pressure.

In the context of paternalistic or algorithmic interactions, if the "gift" is framed too aggressively or if the subsequent demand is perceived as overly coercive, the brain’s response shifts from a social debt to a defensive arousal. In such cases, the drive to re-establish personal sovereignty overrides the evolutionary pressure to reciprocate.

5.1. The Architecture and Measurement of Reactance

To quantify the magnitude of resistance, we propose the implementation of the "Intertwined Model," as developed by Dillard and Shen (2005) and further refined by Quick (2012). Within this framework, reactance is conceptualized not as a single state, but as an inseparable amalgamation of cognitive and affective components.

Specifically, the research program shall utilize the sub-dimensions of "perceived threat to freedom" and "anger" to evaluate the resistance elicited following indebtedness scenarios. By measuring both the affective response (the surge of anger) and the cognitive response (the generation of counter-arguments), the model can delineate the boundary conditions where the "social shackle" of reciprocity breaks under the weight of perceived manipulation.

5.2. Critical Moderators of Reactance

The determination of whether a reciprocal obligation serves as a "weapon of persuasion" or a "catalyst for resistance" is contingent upon three pivotal moderating variables:

- **Source Credibility:** As demonstrated by Allen et al. (2023), individuals integrate the reputations of social partners into their internal reinforcement learning signals. If the benefactor is encoded as "exploitative," computational signals within the Default Network (DN) process the gesture as a strategic "trap." This cognitive appraisal effectively neutralizes the latent sense of indebtedness, preventing the activation of a moral obligation.
- **Indicators of Sincerity (Perceived Effort):** In alignment with Dishop et al. (2025), a low perception of the effort underlying the assistance—frequently observed in the utilization of fully autonomous AI—leads the gesture to be categorized as "insincere." This perception reduces the reciprocal bond to a purely rational "transaction cost," stripping the interaction of its social-emotional weight and fostering cognitive distance rather than loyalty.
- **Transparency and Disclosure:** Explicitly disclosing the agent's identity (human vs. AI) serves a dual purpose. According to the literature on identity disclosure, such transparency reduces the user's sensation of "uncanniness" and the concomitant fear of hidden manipulation. By providing the user with a clear ontological framework for the interaction, transparency significantly mitigates psychological reactance.

5.3. Boundary Conditions: Hyper-personalization and the Surveillance Paradox

In contemporary digital user experience (UX) paradigms, the hyper-personalization trap serves as a primary boundary condition for the principle of reciprocity. When algorithmic systems utilize a user's longitudinal historical context with excessive granularity, they inadvertently trigger psychological reactance by eliciting invasive perceptions of surveillance and persistent monitoring.

Virvou (2023) underscores that mitigating such resistance necessitates the prioritization of design architectures that champion user agency and autonomy. In this framework, the implementation of robust "opt-out" mechanisms and transparent data-governance protocols ensures that reciprocity is recontextualized. By granting the user control over their own data, the platform facilitates a transition where the interaction is interpreted as a consensual social contract rather than a manipulative shackle.

5.4. Empirical Assessment of Reactance: EEG Protocol Proposal

To rigorously quantify the neural responses elicited by violations of fairness and "unsolicited favors," we propose a research design utilizing a modified Ultimatum Game (UG) paradigm, with neural dynamics monitored via Electroencephalography (EEG).

Frontal Midline Theta (FM-Theta): Within the extant literature, Frontal Midline Theta oscillations (4–8 Hz) have been established as robust neural correlates of social pain, cognitive conflict, and perceived transgressions of fairness. In this protocol,

FM-Theta is identified as a primary biomarker of psychological reactance. An escalation in FM-Theta spectral power is anticipated as a direct reflection of dorsal anterior cingulate cortex (dACC) activation, triggered instantaneously by manipulative or inequitable offers from the counterparty.

Protocol Parameters:

- **Electrode Configuration:** In accordance with the International 10-20 system, data acquisition shall focus on the frontal midline regions, specifically electrodes Fz, FCz, and Cz, where conflict-related theta activity is most prominent.
- **Temporal Window:** Analysis shall target the spectral power within the 4–8 Hz band during a 200–500 ms post-stimulus interval, capturing the rapid evaluative phase of the social stimulus.
- **Analytical Framework:** Utilizing the Event-Related Spectral Perturbation (ERSP) method, the model will delineate the neural correlation between the state of perceived indebtedness and the subsequent behavioral act of "offer rejection." This approach allows for the visualization of dynamic changes in the frequency spectrum over time, providing a high-resolution map of the brain’s transition from debt-processing to autonomous resistance.

6. COGNITIVE SELF-DEFENSE AND MITIGATION STRATEGIES

Mitigating the neural pressure inherent in reciprocity and safeguarding individual autonomy requires subjecting this evolutionarily entrenched "error-monitoring" mechanism to a rigorous rational audit. To bolster individual resistance against the strategic manipulation of reciprocity, the following methodological strategies are proposed:

6.1. Strategic Postponement: The 24-Hour Rule

During the phase of acute emotional arousal—specifically at the moment of receiving a significant gift or gesture—the "social conflict" signal within the dorsal anterior cingulate cortex (dACC) reaches its peak intensity. This surge creates a state of cognitive urgency that favors compliance.

The Intervention: Deferring the decision-making process for a minimum of 24 hours facilitates the normalization of the dopaminergic pressure induced by the sudden sense of indebtedness. This temporal buffer enables the recruitment of the prefrontal cortex (PFC), the seat of executive function and rational deliberation.

During this 24-hour interval, the individual is capable of transitioning the "indebtedness entry" from a state of transient transactional processing toward a longitudinal evaluative framework. By allowing the initial affective surge to dissipate,

the user can assess the gesture not as an immediate social debt, but as an isolated data point within a broader utilitarian context.

6.2. Cognitive Relabeling (Neutralizing Indebtedness)

Accurately defining the nature of a perceived gesture is one of the most effective methods for cognitively neutralizing the weight of social obligation. While data encoded as a "gift" triggers an evolutionarily mandated drive for reciprocation, labeling the process as a "strategic intervention" activates psychological reactance and defensive filters.

By consciously altering the semantic category of the stimulus, the individual shifts the interaction from a social exchange framework to a market exchange framework. This transition effectively "mutes" the neural debt signal, as market transactions are governed by contract and utility rather than moral indebtedness.

Reframing Paradigms: To facilitate this cognitive shift, the following relabeling scripts are proposed:

- **From Gift to Tactic:** "This minor concession is not a philanthropic gesture; it is a calculated sales tactic designed to elicit compliance."
- **From Sample to Instrument:** "This complimentary sample functions as a marketing instrument specifically engineered to circumvent my rational deliberative processes."
- **From Personalization to UX Strategy:** "The high-level personalization afforded by this digital assistant represents a UX-driven reciprocity mechanism optimized to facilitate maximal data disclosure."

6.3. AI Filtration and the Inference of Effort

Questioning the provenance of assistance and the magnitude of expended labor serves to attenuate the "invisible shackles" of reciprocity. As empirically substantiated by Dishop et al. (2025), the detection that a service or gesture is AI-generated precipitates a downward revision in effort inference, consequently diminishing the magnitude of felt obligation.

Disclosing or identifying the algorithmic architecture underlying an act of assistance attenuates perceived warmth, thereby facilitating the reduction of social indebtedness to a mere rational transaction cost. In this context, the realization that "the machine does not tire" acts as a profound cognitive liberator; since the agent incurred no biological or psychological cost in providing the benefit, the evolutionary mandate to "repay the favor" is rendered obsolete.

6.4. Mnemonic Awareness and the Regulation of Finite Memory

Research on "finite reward memory," as presented by Glynatsi et al. (2024), indicates that individuals frequently formulate reciprocity decisions based primarily on the preceding two or three dyadic interactions. This suggests that the "social ledger" in the human brain is highly sensitive to recency bias, often overlooking a longer history of transactional imbalance.

From the perspective of cognitive self-defense, the individual can circumvent strategic exploitation by exercising mnemonic regulation. This involves posing a critical, meta-cognitive inquiry: "Is my current sense of indebtedness predicated solely upon a singular, minor gesture in our most recent interaction, rather than the aggregate value of the relationship?" By consciously expanding the temporal window of evaluation, the individual neutralizes the affective salience of the most recent "gift" and restores a rational, longitudinal perspective on the exchange. This prevents the "recency trap" where a single well-timed incentive can trigger a disproportionate reciprocal response.

6.5. The Cognitive Self-Defense Micro-Protocol

To preserve individual autonomy within increasingly complex socio-digital ecosystems, we propose the implementation of the following five-stage micro-protocol. This procedural framework is designed to interrupt the automaticity of the "social shackle" and restore executive control:

- **S.T.O.P. (24-Hour Latency):** Allow for the systematic attenuation of affective arousal before initiating any reciprocal behavior. This temporal buffer ensures that the decision is not driven by the immediate dopaminergic surge of indebtedness.
- **L.A.B.E.L. (Semantic Re-categorization):** Critically evaluate the stimulus. Is the gesture a "philanthropic gift" rooted in genuine prosociality, or is it a "strategic promotional instrument" designed for conversion and compliance?
- **E.F.F.O.R.T. (Agency Inference):** Investigate the provenance of the assistance. Is the source human or synthetic (AI)? Quantify the actual level of personal investment or biological cost incurred by the benefactor.
- **M.E.M.O.R.Y. (Longitudinal Audit):** Counteract the recency bias. Determine whether the current drive to reciprocate is predicated solely upon the transient reward of the most recent 2–3 interactions or the aggregate historical balance of the relationship.
- **B.O.U.N.D.A.R.Y. (Threshold Configuration):** Define an explicit "reciprocal ceiling." Establish pre-determined limits for temporal allocation, financial capital, and emotional labor before engaging in the exchange.

This protocol ensures that reciprocity remains a transparent and equilibrated social contract rather than a manipulative promissory note. By subjecting the evolutionary

"error-monitoring" signals of the dACC to this rational audit, the user effectively reclaims their cognitive sovereignty.

7. CONCLUSION: THEORETICAL CONTRIBUTIONS AND RECOMMENDATIONS

The obligation of reciprocity constitutes a foundational social legacy of human evolution; however, in the contemporary era, it carries the potential to pose a profound threat to individual autonomy. The primary theoretical contribution of this research lies in its empirical demonstration that reciprocity is far more than a mere cultural norm. Rather, it is a neurobiological propensity—governed by the neural nodes of the Anterior Cingulate Cortex (ACC), the Insula, and the Default Network (DN)—that functions to facilitate the homeostatic maintenance of social equilibrium.

While this biological disposition has served as a critical mechanism for the evolution of large-scale social cooperation, it is increasingly susceptible to structural exploitation. In the hands of contemporary digital algorithms and institutional frameworks, the evolutionary pressure to "pay one's debts" can be transmuted into an instrument for constraining individual sovereignty. By weaponizing the brain's internal error-monitoring signals, digital entities can create an "asymmetric social contract" where the user remains in a perpetual state of perceived indebtedness.

Ultimately, this study posits that the survival of autonomy in the digital age depends on our ability to subject these ancient, automated neural impulses to rigorous rational scrutiny. By understanding the neuro-evolutionary architecture of the social shackle, we can begin to design systems that honor—rather than exploit—the human drive for cooperation.

7.1. The AI Dimension and Future Directions: A Research Agenda

The integration of Artificial Intelligence (AI) as a social actor establishes novel inflection points within the dynamics of reciprocity. To further delineate these boundaries, the following research agenda and hypothesis-task configurations are proposed:

- **H1 (Source Manipulation and Indebtedness Decoupling):** AI-mediated assistance generates a significantly lower perception of personal indebtedness compared to human-mediated assistance. This is primarily driven by a diminution in perceived warmth and a higher attribution of "automated ease."
    - **Experimental Paradigm:** Utilizing a design where the source of assistance is manipulated (human vs. AI vs. hybrid models), researchers should quantify scores for warmth, effort inference, and felt obligation via the Stereotype Content Model (SCM) scales.

- **H2 (UX Design and Neuropsychological Loyalty):** As the level of personalization in User Experience (UX) design increases (e.g., nominative address, historical context referencing), the cause-effect reciprocity established through algorithmic "cognitive alleviation" is fortified.
    - **Experimental Paradigm:** In an environment where personalization levels are manipulated (low vs. high), "repayment behaviors"—such as response latency in survey participation and brand loyalty metrics—should be rigorously tested.
- **H3 (Longitudinal Interaction and Social Norm Adherence):** As the duration of AI companionship and interaction proximity increases, users begin to ascribe intentionality and "mind" to agents, leading to the unconscious adherence to social norms such as etiquette and reciprocity.
    - **Experimental Paradigm:** Longitudinal monitoring of interactions with AI companions (e.g., Replika) should be conducted using parasocial bond scales and compliance levels regarding debt norms, representing the "mental shackle" effect over time.
- **H4 (Mnemonic Constraints and Strategic Stability):** Even under conditions of finite reward-memory, iterative digital interactions can transmute reciprocity into an evolutionarily stable strategy (ESS) of exploitation.
    - **Experimental Paradigm:** Under dual-task paradigms designed to heighten cognitive load, the stability of reciprocity and exploitation strategies in digital interactions should be evaluated through game-theoretic modeling.

7.2. Concluding Remarks: The Dual-Edged Sword of the Default Network

The evolutionary expansion of the Default Network (DN) has bequeathed humanity a double-edged sword: while it has cultivated the foundations for deep and complex social cooperation, it has simultaneously enabled sophisticated capacities for exploitation and manipulation. Reciprocity serves as a pivotal mechanism of social regulation that operates on both sides of this divide—functioning as both a liberating force for collective growth and a psychological shackle for individual entrapment. Understanding the neuro-behavioral architecture of this network is not merely an academic exercise; it is a prerequisite for maintaining cognitive sovereignty in the age of algorithmic actors.

REFERENCES

Allen, T. A., Hallquist, M. N., & Dombrovski, A. Y. (2023). The dark side of mentalizing: Learning signals in the default network during social exchanges support cooperation and exploitation. *bioRxiv*. https://doi.org/10.1101/2023.05.03.538867

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions.* HarperCollins.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*(1), 124–140.

Arslan, M. (2012). İş ahlakı açısından Türk iş kültüründe paternalizm ve kurumsallaşma [Paternalism and institutionalization in Turkish business culture in terms of business ethics]. *Turkish Journal of Business Ethics, 5*(9), 115–142.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652.

Brehm, J. W. (1966). *A theory of psychological reactance.* Academic Press.

Buckner, R. L., & DiNicola, L. M. (2019). The organization of the human cerebral cortex estimated from intrinsic functional connectivity. *PNAS, 116*(51), 25374–25376.

Chaturvedi, R., Verma, S., & Srivastava, V. (2024). Empowering AI companions for enhanced relationship marketing. *California Management Review, 66*, 65–90.

Cialdini, R. B. (2021). *Influence, new and expanded: The psychology of persuasion.* Harper Business.

Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*(1), 59–70.

Dillard, J. P., & Shen, L. (2005). On the nature of reactant rhetoric: Using helical structures to explore the mechanism of resistance to persuasion. *Human Communication Research, 31*(2), 144–168.

Dishop, C. R., Brown, A. S., Chao, P. Y., Kuznetsov, A., & Woolley, A. W. (2025). Machines in the middle: Using artificial intelligence (AI) while offering help affects warmth, felt obligations, and reciprocity. *Journal of Business and Psychology.* https://doi.org/10.1007/s10869-025-10068-x

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 498–541). John Wiley & Sons.

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review, 96*(5), 1611–1630.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Flynn, F. J. (2003). What have you done for me lately? Temporal adjustments to favor evaluations. *Journal of Applied Psychology, 88*(3), 540–545.

Glynatsi, N. E., McAvoy, A., & Hilbe, C. (2024). Evolution of reciprocity with limited payoff memory. *Proceedings of the Royal Society B, 291*(20232493).

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review, 25*(2), 161–178.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619–619.

Krauter, S. (2024). *Synthetic souls and social debt: The evolution of reciprocity in human-AI systems.* [Academic Monograph/In Press].

Łukasik, A., & Gut, A. (2025). From robots to chatbots: Unveiling the dynamics of human-AI interaction. *Frontiers in Psychology, 16*, 1569277. https://doi.org/10.3389/fpsyg.2025.1569277

Mauss, M. (1925). *The gift: Forms and functions of exchange in archaic societies.* Routledge.

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 Tesla, with 16-fold acceleration using partial parallel imaging with SNR efficient simultaneous imaging. *Magnetic Resonance in Medicine, 63*(5), 1144–1153.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*(7063), 1291–1298.

Özkarar-Gradwohl, F. G. (2019). Cross-cultural affective neuroscience. *Frontiers in Psychology, 10*, 794. https://doi.org/10.3389/fpsyg.2019.00794

Palmatier, R. W., Jarvis, C. B., Bechkoff, J. R., & Kardes, F. R. (2009). The role of customer gratitude in relationship marketing. *Journal of Marketing, 73*(5), 1–18.

Quick, B. L. (2012). The intertwined model of reactance. In J. P. Dillard & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 165–183). SAGE Publications.

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433–447.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395–405.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron, 79*(2), 217–240.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* John Wiley & Sons.

Singer, T., Seymour, B., O'Doherty, J., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466–469.

Strohmetz, D. B., Rind, B., Fisher, R., & Lynn, M. (2002). Sweetening the till: The use of candy to increase restaurant tipping. *Journal of Applied Social Psychology, 32*(2), 300–309.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry. *Psychological Science, 19*(4), 339–347.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*(1), 35–57.

Uğurbil, K., et al. (2013). Pushing the limits of fMRI: High resolution at 7T. *NeuroImage, 80*, 80–104.

Usta, S., Selcuk, B., & Okcay, H. (2024). Majority members' acculturation: The role of social norms and perceived warmth in the Turkish context. *Current Psychology.* https://doi.org/10.1007/s12144-024-07011-3

Virvou, M. (2023). Artificial Intelligence and User Experience in reciprocity: Contributions and state of the art. *Intelligent Decision Technologies, 17*, 73–125. https://doi.org/10.3233/IDT-230092

Zhang, Y., Chen, R., Liu, S., Chen, P., & Mai, X. (2025). Neural correlates of reciprocity bias: Social debts modulate unfairness perception of violation during third-party observation. *Cerebral Cortex, 35*(2), bhae497. https://doi.org/10.1093/cercor/bhae497

DECLARATION

Author’s Declaration of Originality I hereby certify that this paper, titled *"Invisible Shackles of the Mind and the Anatomy of Social Indebtedness: A Neuropsychological and Cultural Analysis of the Principle of Reciprocity”*, is my original work. All sources and intellectual contributions have been appropriately cited. This document has been digitally signed to verify its authenticity and the integrity of the data presented herein. System ID: Ahmet.dll / Protocol: X-09